Lab 1 Requirements

1. Team Members

- Călăuz Răzvan
- Coșofreț Octavian

2. Team Name

- FruitGuardians

3. Team Members Roles

- Călăuz Răzvan: Model Architect, Evaluation Specialist, ML Engineer
- Coșofreț Octavian: Data Engineer, ML Engineer

4. Project Theme

- "Detection of Diseased Fruits for the Purpose of Being Harvested Faster or for Applying an Optimal Treatment."

5. Project Name

- HarvestGnosis

6. Relevant Research Articles

1. "Adapted Approach for Fruit Disease Identification using Images": https://arxiv.org/abs/1405.4930
    a. Dataset: A custom dataset which comprises four different categories: Apple Blotch (104), Apple rot (107), Apple scab (100), and Normal Apple (80). The total number of apple fruit images (N) is 391.
    b. Algorithms Used: K-means clustering for segmenting defects in images. The classification is done using a standard SVM multiclass (no parameters mentioned).
    c. Metrics: accuracy 93%.
    d. Results: The K-means approach can identify the apple diseases correctly with an accuracy of 93%.

2. "Disease Detection in Fruits using Image Processing": [Disease Detection in Fruits using Image Processing | IEEE Conference Publication | IEEE Xplore]
    a. Dataset: The images of fruits of reddish fruits presenting both healthy and with a certain disease were gathered manually, using phone cameras from different agricultural regions. The pictures were taken from different angles, in different lighting conditions and saved as JPGs. It seems that the acquired data was supplemented using an unnamed dataset from Kaggle.com. The preprocessing part of data included rotating images such that all are in a vertical position, reshaping to a 250 X 250 pixels and increasing image sharpness. The image segmentation process stated in the paper involved converting pre-processed images to multiple color spaces(L*a*b, HSV, Grey, RGB) before transforming them in a binary format which will be used in the convolution part
    b. Algorithms Used: The main focus of this paper was using Convolutional Neural Networks(CNN) as primary learning algorithm although not much implementation details are given on this.
    c. Metrics: The metrics evaluation of the paper are limited, focusing primarily on accuracy indicators. Based on the only indicators used, the model achieved an overall accuracy of 97% for the CNN-based classification system.
    d. Results: The proposed solution successfully identified and classified three distinct fruit diseases: Bitter rot, Powdery mildew, Sooty blotch, highlighting the disease-affected areas on the fruit images. No time-specific results are provided, although it claims that the results are provided in "few seconds". No comparison with existing approaches to provide context for the 97% accuracy.

3. "Fruit Disease Detection and Classification using Machine Learning and Deep Learning Techniques": https://www.ijisae.org/index.php/IJISAE/article/view/3805
    a. Dataset: The dataset used contains different varieties of fruits, both collected with help of drones, sensors and cameras, but also using the Citrus and PlantVillage datasets. No image counts are

provided. Preprocessing of data starts with cropping the images to isolate relevant fruit regions followed by applying filters to reduce noise and increase contrast. Color space transformations were performed from RGB to Grey scaled. Different image segmentations were done before the feature extraction: edge-based, region-based, threshold-based, and clustering-based to isolate diseased regions from healthy fruit tissue. The feature extraction process involved three primary categories: color feature vectors, shape feature vectors, and texture feature vectors, which were extracted after segmentation to characterize both healthy and diseased regions of the fruits. Specific feature extraction techniques mentioned include Global Color Histogram, Color Coherence Vector, Local Binary Pattern (LBP), and Complete LBP

b. Algorithms used: The research employed a hybrid CNN (HCNN) architecture that combined VGGNET-16 with YOLOV8 segmentation capabilities. The algorithms received multi-dimensional input from a hierarchical preprocessing pipeline where segmented fruit images yielded color, texture, and shape-based feature vectors that characterized disease manifestations across various fruit types. The proposed model demonstrates performance advantages, but lacks implementation details, like hyperparameter optimization or training procedures.
c. Metrics: Primary metrics used were accuracy(97.10%), precision(0.97) and recall(0.98), demonstrating the hybrid model as being a strong approach for fruit disease classification.
d. Results: The results demonstrate that the integration of specialized segmentation capabilities (via YOLOV8) with the feature extraction power of established CNN architectures (VGGNET-16) creates synergistic performance improvements. The comparative analysis evaluated 14distinct classification algorithms including traditional approaches (ANN, SVM, Naive Bayes, Random Forest, Fuzzy) and deep learning methods (PNN, RNN, RNN+LSTM, CNN, FCNN, DCNN, DNN, GAN), with the proposed hybrid model demonstrating optimal precision (0.97) and recall (0.98).

4. "An AI-based Framework for Disease Recognition in Fruit Leaf using Deep Learning Methods": https://www.researchgate.net/publication/389495019_An_AI_-based_Framework_for_Disease_Recognition_in_Fruit_Leaf_using_Deep_Learning_Methods
    a. Dataset: The dataset has two categories (apples and grapes leaf). No dataset name mentioned. There are four data classes in the Apple dataset: one class of healthy leaves and three classes of diseased leaves. Cedar rust, scab, and black rot are the other three classes. There is the same amount of images in each of the four classes, or 6000 total. Four groups of data are included in the grape dataset: healthy (1446 images), black rot (2960 images), esca (black measles) (3400 images), and leaf blight (2800 images).
    b. Algorithms Used: The framework uses a pre-trained DCNN (Deep Convolutional Neural Network) that is fine-tuned on both the original and noisy data. Features are extracted (from the average pooling layers) and then fused using serial concatenation. Optimization - PFA (Path Finder Algorithm). There was also used ANN classifier, but DCNN has shown better results. Hyperparameters' values aren't listed, but they are optimized to maximize accuracy.
    c. Metrics: DCNN on Apple Dataset: Accuracy (97.75%), Precision (96.67%), Recall (93.78%), F-Score (93.94%). DCNN on Grape Dataset: Accuracy (97.45%), Precision (98.87%), Recall (89.65%), F-Score (82.44%).
    d. Results: The proposed model outperforms several other recent approaches. Badjie and E. D. Ülker accuracy - 93.83%, I Haider, M.A Khan, M Nazir, TKimAnd Jae-Hyuk Cha accuracy - 95.65%, proposed model accuracy - 97.40%.